renumbering

Claim Amendments

Claim 1. (currently amended) A computerized storage and retrieval system of biological information comprising:

a means for entering data [[entry]];

a means for displaying the data;

a programmable central processing unit ~~for performing~~ (programmed to perform) automated analysis using semi-linear pathway representation (SLIPR) and methods to compare and predict pathways and their components and to establish a pathways database; [[and]]

a ~~data storage~~ means for storing data; and

a relational database containing annotated and curated ~~protein~~ pathways and ~~annotated~~ information on the pathways and their components ~~stored in a relational database, wherein the pathways are annotated and organized in a curated clustering arrangement and the annotated information is accessed through the relational database~~.

Claim 2. (currently amended) The system of claim 1, wherein the information pertaining to the pathways and their components and [[is]] stored in a plurality of tables further comprises ~~comprising~~ proteins, their sequences and attributes; protein interactions; protein-protein associations; protein pathways; mRNA, microarray, and protein expression data; genes, their sequences and attributes; [[and]] descriptions of cells, tissues, and organs[[,]]; pathology reports [[,]]; and patient histories [[,]] and treatments.

Claim 3. (currently amended) The system of claim 1, wherein the central processing unit is (further) programmed to run an algorithm or a method for constrained clustering or an optimization algorithm for dynamic programming and one or more functions or standard methods selected from retrieve, input, edit, annotate, search, calculate similarities, align, compare, optimize using linear next-neighbor criteria or global minimization criteria, and predict homologous or orthologous ~~protein~~ pathways and their components and wherein the method for constrained clustering is further selected from average linkage, single linkage, complete linkage, K-means, or self-organizing maps.

Claim 4. (6) (currently amended) The system of claim 1, wherein the central processing unit is further programmed to perform one or more analyses selected from protein sequence analysis, protein interactions analysis, protein-protein association analysis, protein pathway analysis, gene expression analysis, pathway annotation analysis, pathway edit analysis, pathway expression analysis, tissue expression analysis, subtractive hybridization analysis, electronic northern analysis, and [[or]] commonality analysis.

Claim 5. (8) (currently amended) The system of claim 1, wherein each ~~the data is entered using the standard for~~ pathway in the database further comprises at least two nodes connected by a mode ~~representation~~.

Claim 6. (currently amended) The system of claim 5 [[1]], wherein ~~a means for displaying the data is used to show two related pathways as a diagram containing nodes which represent proteins or non-protein molecules; nodes that represent protein interactions or protein-protein associations;~~ at least two pathways are compared using scores calculated from sequence identity, motif or structural homologies that interrelate nodes [[;]] and employing coefficients of similarity calculated or selected from gene expression including clone counting of cDNAs, ESTs, promoter regions, or regulatory sequences; electronic northern analysis, immunoprecipitation, microarray data, quantitative real-time polymerase chain reaction, transcript images, and yeast two hybrid data; a promoter similarity matrix, a protein-protein association matrix and a similarities among protein interactions matrix (SCIM matrix) that interrelate modes of the pathways.

Claim 10. (currently amended) The system of claim 1, wherein the central processing unit is further programmed to ~~compare~~ perform pairwise comparison of at least two ~~protein~~ pathways or their components selected from [[by a]] node-only, [[a]] mode-only, and [[or a]] node-and-mode comparison and wherein the node-only comparison is further selected from protein only, non-protein only, and protein and non-protein nodes.

Claim 4. (currently amended) The system of claim 3 [[1]], wherein the central processing unit ~~is programmed to run~~ runs the [[an]] optimization algorithm for dynamic programming further comprising:

a) initializing an array, in which a two dimensional array $M=M_{ij}$ with J rows of nodes and variant length for each row, the length for i-th row is n_i is set up and $M_{ji}=0$, where $1<=i<=n_J$,

b) backfilling the array via backward recursion with the formula

$$M_{ik} = \max_{\substack{j\neq i \\ 1\le l\le n_j}} \left\{ w\left(a_{ik}, a_{jl}\right) + M_{jl}\theta\left(w\left(a_{ik}, a_{jl}\right)\right) \right\} \text{ for } 1 \le k \le n_i,\ 1 \le i \le J$$

where è(.) is the step function defined as è (v)={0, if v<=0; 1, if v>0} and w(.,.) is the scoring function between the two nodes, defined as

$$w(a_{ik}, a_{jl}) = \begin{cases} 0, \text{ if } i = j,\ a_{ik} = a_{jl},\ a_{ik} = -D, \text{ or } a_{jl} = -D \\ \text{where } D > 0, \text{ or} \\ \theta\left[c_{ik,jl} - t_c\right] \bullet \left\{ \alpha\left(1 - \left|s_{ik} - s_{jl}\right|\right) + \left(1 - \alpha\right) c_{ik,jl} \right\} \text{ otherwise} \end{cases} ; \text{ and}$$

c) ~~using traceback to identify~~ finding the best alignment to predict nodes of putative pathways PPW_j, $1<=j<=\max n_i$ with the top *n* best scores.

Claims 9-22. (canceled)

Claim 23. (currently amended) The system of claim 3 [[1]], wherein the central processing unit ~~is programmed to run a~~ runs the constrained clustering algorithm further comprising:

a) assigning a distance between a protein in a query pathway and each protein ~~every pair of proteins~~ in the database;

b) merging at each round the [[two]] protein in the query pathway and the protein in the database with the closest distance [[pairs]] into a cluster until user-set threshold is met; and

c) computing mean linkage between two clusters using

$$d(c_1, c_2) = \left\{ \frac{1}{n_1 + n_2} \sum_{i \in c_1} \sum_{j \in c_2} d_{ij}^{\rho} \right\}^{1/\rho}$$

where~~[[is]]~~ n_1 is the size of c_1 and n_2 is the size of c_2 and the distance between the two clusters is a weighted average of the distance between two proteins i and j and a function of similarity to query protein ($s_i = S_{iq(i)}$, $s_j = S_{jq(j)}$) where q(i) is the query protein homologous to database protein i and protein-protein association (a_{ij}) and

$$d_{ij} = \alpha(1 - a_{ij}) + (1 - \alpha)|s_i - s_j|$$

Claim 24. (canceled)

Claim 25. (currently amended) The system of claim 7, wherein ~~pathway analysis uses a~~ node-and-mode comparison predicts new homologous or orthologous pathways by ~~comprising:~~

~~a) submitting a query pathway and protein sequences;~~

~~b)~~ comparing nodes using an optimization algorithm for dynamic programming or a constrained clustering algorithm or standard methods and ~~wherein a sequence identity score or p-value summarizes similarity and a weighting factor between 0 and 1 is assigned to corresponding nodes;~~

~~c) comparing~~ modes by generating a SCIM matrix ~~wherein a~~ containing coefficients of similarity ~~is assigned to corresponding modes;~~

~~d) aligning pathways globally or locally, wherein insertion or deletion of nodes or modes incurs a penalty;~~

~~e) summing all similarity scores; and~~

~~f) displaying at least one high-scoring segment of the aligned pathways derived from node and mode comparison, thereby performing pathway analysis~~.

Claim 26. (currently amended) A process for ~~The system of claim 1, wherein the central processing unit is programmed to perform~~ analyzing or predicting novel homologous or orthologous pathways ~~analysis using a submitted query pathway and protein sequences~~ comprising:

a) submitting a query pathway and protein sequences; to the system of claim 1

b) organizing the pathway and sequences using SLIPR ~~the standard for pathway representation;~~

c) [[b)]] comparing each sequence ~~protein sequences~~ of the query pathway with each sequence ~~all protein sequences~~ in the pathways database using an algorithm for constrained clustering or an optimization algorithm for dynamic programming and standard methods of protein comparison to calculate sequence identity scores;

d) [[c)]] using a SCIM matrix derived from protein interactions or protein-protein association analysis to compare coefficients of similarity for each interaction of the query pathway and each interaction ~~all interactions for proteins~~ in a pathway in the pathways database;

e) [[d)]] calculating an overall similarity score (OS-score) based on sequence identity scores and coefficients of similarity for each pathway; and

~~e) removing all pathways not meeting a user-specified threshold for OS-score; and~~

f) retrieving ~~aligned~~ pathways meeting a user-specified [[the]] threshold for OS-score, thereby ~~performing~~ analyzing or predicting novel homologous or orthologous pathways ~~analysis~~.

Claim 27. (currently amended) The system of claim 4 [[3]], wherein ~~pathways database is searched for protein interactions comprising:~~

~~a) submitting a query pathway;~~

~~b) performing~~ protein interactions analysis between a [[the]] query pathway and each pathway [[all]] ~~pathways~~ in the ~~pathways~~ database produces a ~~wherein~~ coefficient of similarity ~~is produced~~ to interrelate each mode of the query pathway and a mode of the most closely related protein pathway [[;]]

~~c) retrieving at least one pathway alignment; and~~

~~d) showing the alignment based on the search for protein interactions.~~

Claim 28. (canceled)

Claim 29. (currently amended) The process of claim 26, wherein ~~The system of claim 3, wherein a query pathway to search a pathways database to predict~~ the predicted orthologous ~~pathways comprising:~~

~~a) submitting a query pathway and known protein sequences;~~

~~b) comparing known sequences to all protein sequences stored in the database;~~

~~c) retrieving~~ pathway further comprises orthologous proteins with the highest sequence identity to the ~~known~~ proteins in the query pathway [[;]] and

~~d) inheriting~~ protein interactions inherited from the query pathway[[; and]]

~~e) aligning the query pathway and the orthologous proteins, thereby predicting orthologous pathways.~~

Claims 30-36. (canceled)